TREATMENT OF MODEL ERROR IN ENSEMBLE DATA ASSIMILATIONS DUE TO UNREPRESENTED SCALES OF MOTION

Thomas M. Hamill[1] and Jeffrey S. Whitaker[2]

[1]*University of Colorado and NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado*

[2]*NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado*

29 May 2004

submitted to *Monthly Weather Review*

DRAFT

Corresponding author address: Dr. Thomas M. Hamill, NOAA-CIRES CDC, R/CDC 1, 325 Broadway, Boulder, CO 80305-3328. e-mail: tom.hamill@noaa.gov, ph: (303) 497-3060

ABSTRACT

Insufficient model resolution is one source of model error in numerical weather predictions. Methods for parameterizing this error in ensemble data assimilations were explored here. Experiments were conducted with a 2-layer primitive equation model, where the assumed true state was a T127 forecast simulation. Ensemble data assimilations were performed with the same model at T31 resolution, assimilating imperfect observations drawn from the T127 forecast. The magnitude of errors due to model truncation was much larger than the error growth due to initial condition uncertainty, making this a severe test of the ability of an ensemble-based data assimilation to deal with model error. Two general methods, "covariance inflation" and "additive error," were considered for parameterizing the model error at the resolved scales (T31 and larger) due to interaction with the unresolved scales (T32 to T127). Covariance inflation inflates the background forecast members' deviations about the ensemble mean, while additive error adds specially structured noise to each ensemble member forecast before the update step.

The method of parameterizing this model error has a substantial effect on the accuracy of ensemble data assimilations. Covariance inflation produced ensembles with analysis errors that were no lower than the analysis errors from 3-dimensional variational (3D-Var) assimilation, and for the method to avoid filter divergence, the assimilations had to be periodically reseeded. Actual model errors depended on the dynamics, growing more in the middle latitudes, while covariance inflation uniformly inflated the model spread. This caused a progressive degradation of the ability of the ensemble to span the actual forecast error. The most accurate model error parameterization was an additive model error parameterization, which reduced the error difference between 3D-Var and a near-perfect assimilation system by $\sim$ 40%. In the lowest-error simulations, additive errors were parameterized using samples of model error from a time series of differences between T63 and T31 forecasts. Scaled samples of differences between model forecast states separated by 24 h were also tested, as well as scaled samples of the model state's anomaly from the model climatology. These latter two methods of generating additive error samples pro-

duced analyses that were progressively less accurate. The decrease in accuracy was likely due to their inappropriately long spatial correlation length scales.

1. INTRODUCTION

Ensemble-based atmospheric data assimilation techniques are actively being explored as a potential replacement for or complement to 3- or 4-dimensional variational data assimilation (3D-Var or 4D-Var; Parrish and Derber 1992, Le Dimet and Talagrand 1986, Courtier et al. 1994, Rabier et al. 1998, 2000). Data assimilation algorithms statistically adjust prior forecasts to newly available observations to generate reduced error initial conditions suitable for numerical weather predictions. As such, these methods require error statistics of both the prior forecast (the "background") and the observations. Typically, error distributions are also assumed to be normally distributed, stationary (time invariant), and perhaps homogeneous (spatially invariant in some sense, such as along a latitude circle). Ensemble-based data assimilation methods relax some of these assumptions. In particular, the background-error statistics are estimated from an ensemble of forecasts and can vary in magnitude and spatial structure depending on the flow of the day. This permits them in theory to provide a more accurate adjustment of forecasts to new observations, resulting in reduced-error analyses. Ensemble-based techniques typically require an ensemble of ~ 20 to a few hundred members of short-range forecasts and as many parallel data assimilation updates. Hence, these methods are computationally expensive, similar to 4D-Var. For more background on ensemble-based methods see, for example, Evensen (1994), Evensen and van Leeuwen (1996), Houtekamer and Mitchell (1998), Burgers et al. (1998), Tippett et al. (2003), Anderson (2003), Evensen (2003), and Lorenc (2003).

In perfect-model tests, the ensemble filters produced simulations with dramatically lower errors than were achieved with competing methods (e.g., Hamill and Snyder 2000, 2002, Anderson 2001). Only within the last year or two have ensemble-based methods been tested in realistic numerical weather prediction models with real observations. In such a case the perfect-model assumptions must be dropped, and the forecast uncertainty due to model error must be parameterized in some fashion. Houtekamer et al. (2004) propose parameterizing model deficiencies with "additive error," adding noise to each member of the ensemble of background forecasts, in their case selecting noise consistent in structure with a 3D-Var background-error covariance

model. Their initial results showed that ensemble-based assimilation methods were competitive with but not superior to 3D-Var. Whitaker et al. (2004, hereafter W04) tested the assimilation of a sparse network of surface pressure observations for purposes of generating a long-term, historical reanalysis. W04 found that the analysis errors from the ensemble data assimilation were significantly less than were achieved with a simple 3D-Var. With the sparse observations, the form of the background-error covariance model had more of an impact on the accuracy of the assimilations (Hamill and Snyder 2000). In W04, model errors were treated through "covariance inflation," inflating the ensemble member's spread about the ensemble mean. Despite the generally good results, the data assimilations produced too little analysis spread in data-rich regions and too much spread in data-sparse regions.

How important is the specific method of parameterizing model errors in ensemble-based data assimilation applications? This is a question we seek to answer in part in this manuscript. Model error, of course, can be introduced from many causes, such as imperfect parameterizations (Buizza et al. 1999, Palmer 2001). We choose to examine only one particular aspect of model error, the errors introduced by the truncation of the forecast model and the resulting lack of interaction with the smaller scales of motion. In these simulations the lack of interaction results in a deficiency of spread but does not systematically bias the ensemble very much. We then consider whether any of several types of additive noise can be as good or better a parameterization of this model error than covariance inflation. Our simulation experiment will assume that the true state is a global simulation at triangular truncation T127, while the assimilation will be carried out at T31 resolution, assimilating imperfect observations sampled from the T127 nature run. We will test whether differences between T63 and T31 forecasts are an effective additive model error parameterization, and we also test the effects of scaled differences between forecast model states separated by 24 h and scaled differences of random model states from the climatological mean state.

Because of its computational speed and general resemblance to operational numerical weather prediction (NWP) models, we will use a simple, global 2-layer primitive equation (PE) model

for these ensemble data assimilation experiments. We begin with an examination of the characteristics of the model itself and of short-term forecast errors due to truncation (section 2). Section 3 provides a description of the ensemble-based data assimilation methodology and various candidate techniques for paramerizing model error. Section 4 describes the experiment, section 5 examines the relative accuracy of the assimilations using various model-error parameterizations and discusses these results, and section 6 concludes.

2. FORECAST MODEL

a. Model design

Results in the rest of the paper will be based on experiments with a dry, global, two-layer PE model. The forecast model was described in Zou et al. (1993, their appendix A) and was used in Hamill et al. (2001) for ensemble data assimilation experiments in a perfect-model context. The model is spectral and was run at three triangular truncations, T127, T63, and T31. The model state vector consists of vorticity and divergence spectral coefficients at two levels as well as coefficients of layer thickness $\Delta\pi$, where π is the Exner function. There is a simple, zonal wavenumber 2 terrain with a maximum amplitude of 2000 m at 45^o N and S latitude and 45^o E longitude and 135^o W longitude, tapering to 0 m at the poles and equator. The model is forced by Newtonian relaxation to a prescribed interface Exner function with a damping timescale τ_{diab} = 20 days. A fourth-order Runge-Kutta scheme is used for the numerical integration. The timesteps are 5, 10, and 20 minutes for the T127, T63, and T31 resolutions. There is $\nabla^8 = (\nabla^2)^4$ hyperdiffusion with a 6-hour e-folding timescale for the shortest resolvable scale. The diffusion acting on a given wavenumber will therefore increase as the truncation is made more severe. Other parameters are the same as in Zou et al. (1993), with the exception that the upper-layer potential temperature is specified to be θ_2=310K, θ_1=280K, the model top $z_{top} = 1.5 \times 10^4\ m$, and the lower layer drag coefficient τ_{drag} = 4 days. The Exner function at the model top $\pi_{top} = c_p - \frac{g z_{top}}{\theta_1}$ is fixed and set to a value of $\sim 478.5 Jkg^{-1}K^{-1}$. The surface Exner function is diagnosed from $\pi_top + \Delta\pi_2 + \Delta\pi_1$.

In the experiments that follow, the model is initialized from a state of rest plus a barotropic vorticity perturbation and allowed to spin up for 100 days before beginning the forecast and assimilation experiments.

b. Model and truncation error characteristics

Error doubling times, as computed from the leading Lyapunov vectors (Legras and Vautard 1996), are faster when the model is run at higher resolution. The error doubling times is 3.78 days at T31, 2.16 days at T63, and 1.88 days at T127. As in Snyder and Hamill (2003), the leading Lyapunov vectors are closely related to the jet-stream dynamics (not shown).

For all resolutions, the model exhibits westerly jets in the middle latitudes and an easterly jet in the tropics (Fig. 1a). The tropical easterly jet is less pronounced at T127 resolution while the mid-latitude westerly jet is more pronounced. The mid-latitudinal gradient of the upper-layer thickness $\Delta\pi$ (Fig. 1c) is slightly larger in magnitude in the T127 simulation, consistent with the stronger mid-latitude jet.

Figure 2 shows the kinetic-energy power spectra from nature runs at T127, T63, and T31 resolutions as a function of total wavenumber k. The spectral slope at subsynoptic scales is shallower than k^{-3} expected with 2-dimensional turbulence but steeper than $k^{-5/3}$ of 3-D turbulence. At wavenumbers around 30, the power in the T31 nature run is damped relative to the two higher resolution simulations, a consequence of the ∇^8 hyperdiffusion which selectively damps the shortest retained scales.

For subsequent experiments, let us assume that our forecast system is only able to resolve scales T31 and larger, hereafter called the "resolved scales." The errors in the resolved scales due to interaction with the scales of motion smaller than T31, the "unresolved scales," is what constitutes the model error. We will assume that the "real" atmosphere evolves according to the forecast dynamics described by the same model, but at T127 resolution. Short-term model error is diagnosed as follows: first, let $\mathbf{x}^{127}(t)$ denote the true model state at time t at full T127 resolution. Let $\mathcal{M}^{127}(\,\cdot\,)$ denote the forecast model operator at T127 resolution taking the state forward one time unit: $\mathbf{x}^{127}(t+1) = \mathcal{M}^{127}(\mathbf{x}^{127}(t))$. Let $\mathcal{T}[\,\cdot\,]$ denote the truncation operation of

the model state to the resolved scales. Then short-term model error is thus $\mathcal{M}^{31}(\mathcal{T}[\mathbf{x}^{127}(t)]) - \mathcal{T}[\mathbf{x}^{127}(t+1)]$. That is, model error consists of the difference of a low-resolution forecast from a truncated initial condition minus the truncated forecast from a high-resolution model using a high-resolution initial condition. This describes the short-term forecast error in the resolved scales due to the lack of modeling of interaction with the unresolved scales.

Figure 3 shows a time series of model errors accumulated independently each 24 h over 5 consecutive time periods. At the beginning of each 24-h period, the T31 simulation was reinitialized with the truncated T127 nature run's state, and model error was accumulated over the next 24-h period. The model error appears to have some temporally correlated features associated with certain short waves. There was also a substantial amount of error that appeared to be more random. Perturbations were generally larger in the regions of large gradients of $\Delta\pi$, typically in the midlatitudes.

Model errors were initially small in scale but grow upscale (Fig. 4), peaking eventually at the synoptic scales. After an initial transient period, these results take on many of the characteristics of the predominantly baroclinic growth witnessed in Tribbia and Baumhefner (2004) rather than the classic upscale growth associated with a k^{-3} spectrum discussed in Lorenz (1969) and Leith and Kraichnan (1972). However, the errors here do not accumulate as quickly at the planetary scales as in the Tribbia and Baumhefner study.

Figure 4 also shows model error spectra when the model was truncated at T63 resolution instead of T31. The overall T63 24-h model error spectrum was approximately an order of magnitude smaller, since far more of the power of the overall spectrum was resolved. The peak power was at a slightly shorter wavenumber, and the errors grew faster, so the T63 model errors were a greater fraction of the T31 model error at 120 h than at 3h. Errors at the smallest scales of the T63 simulation appear to be supersaturated at 3h (a variance higher than climatological variance) but were subsequently damped by the diffusion. Overall, the amplitude of model error due to truncation is decreased as resolution is increased and as the time between assimilation cycles is decreased.

3. ENSEMBLE DATA ASSIMILATION METHODOLOGY

The general ensemble-based assimilation methodology proceeds as follows: First, assume that a set of n perturbed initial conditions is available that presumably samples from the distribution of background (first guess) errors. Perform n + 1 parallel data assimilation cycles using an ensemble-based assimilation algorithm, updating the ensemble mean and perturbations from the mean to the newly available observations, modeling the background-error covariances using the ensemble. Next, make n forecasts forward to the next data assimilation time. This step may include adjusting these forecast ensemble members in some manner to account for model errors. Repeat the update and forecast steps. Below, we consider the details of the update and forecast steps.

a. Updating with the ensemble square-root filter

The assimilation scheme used here has been named the *ensemble square-root filter*, or "EnSRF." A complete description of it and the rationale for its use is provided in Whitaker and Hamill (2002). The underlying principle is to run an ensemble of parallel forecast and data assimilation cycles, ensuring that the ensemble mean analysis and the analysis-error covariance as estimated by the ensemble are consistent with those predicted by Kalman-filter theory.

Let $\mathbf{y}^{\mathrm{o}}(t)$ be a set of observations at time t and $\mathbf{H}$ be a linear operator that converts the model state to the observation space. Let $\mathbf{X}^{\mathrm{b}}(t) = [\ \mathbf{x}^{\mathrm{b}}_1(t), \ldots, \mathbf{x}^{\mathrm{b}}_n(t)\ ]$ be a matrix whose column vectors are the n ensemble of forecasts, and similarly, let $\overline{\mathbf{X}^{\mathrm{b}}}(t) = [\ \overline{\mathbf{x}^{\mathrm{b}}}(t), \ldots, \overline{\mathbf{x}^{\mathrm{b}}}(t)\ ]$ denote an matrix where each column vector is the ensemble mean forecast: $\overline{\mathbf{x}^{\mathrm{b}}}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}^{\mathrm{b}}_i(t)$. Let $\mathbf{P}^{\mathrm{b}}(t)$ be the background-error covariance matrix approximated using the sample covariance from an ensemble of model forecasts; $\mathbf{P}^{\mathrm{b}}(t) = \frac{1}{n-1}\ [\ \mathbf{X}^{\mathrm{b}}(t) - \overline{\mathbf{X}^{\mathrm{b}}}(t)\ ]\ [\ \mathbf{X}^{\mathrm{b}}(t) - \overline{\mathbf{X}^{\mathrm{b}}}(t)\ ]^{\mathrm{T}}$. Let $\mathbf{R}$ be the observation-error covariance matrix.

Following Whitaker and Hamill (2002), it is convenient in the EnSRF to update the equations for the ensemble mean (denoted by an overbar) and each member's deviation from the

mean (prime) separately:

$$\overline{\mathbf{x}}^{\mathrm{a}}(t) = \overline{\mathbf{x}}^{\mathrm{b}}(t) + \mathbf{K}[\ \mathbf{y}^{\mathrm{o}}(t) - \mathbf{H}\overline{\mathbf{x}}^{\mathrm{b}}(t)\ ], \tag{1}$$

$$\mathbf{x}'^{\mathrm{a}}_{i}(t) = (\mathbf{I} - \widetilde{\mathbf{K}}\mathbf{H})\ \mathbf{x}'^{\mathrm{b}}_{i}(t). \tag{2}$$

Here, the superscript ^{a} denotes the analysis, $\mathbf{K}$ is the traditional Kalman gain,

$$\mathbf{K} = \mathbf{P}^{\mathrm{b}}(t)\mathbf{H}^{\mathrm{T}}\ [\mathbf{H}\mathbf{P}^{\mathrm{b}}(t)\mathbf{H}^{\mathrm{T}} + \mathbf{R}]^{-1}. \tag{3}$$

and $\widetilde{\mathbf{K}}$ is a the "reduced" gain used to update deviations from the ensemble mean. When sequentially processing independent observations, $\mathbf{K}$, $\widetilde{\mathbf{K}}$, $\mathbf{H}\mathbf{P}^{\mathrm{b}}(t)$ and $\mathbf{P}^{\mathrm{b}}(t)\mathbf{H}^{\mathrm{T}}$ are all vectors with the same number of elements as the model state vector, and $\mathbf{H}\mathbf{P}^{\mathrm{b}}(t)\mathbf{H}^{\mathrm{T}}$ and $\mathbf{R}$ are scalars. Thus, as first noted by Potter (1964),

$$\widetilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathbf{H}\mathbf{P}^{\mathrm{b}}(\mathrm{t})\mathbf{H}^{\mathrm{T}} + \mathbf{R}}}\right)^{-1} \mathbf{K}. \tag{4}$$

The quantity multiplying $\mathbf{K}$ in Eq. (4) is a scalar between 0 and 1. This means that, in order to obtain the analysis-error covariance consistent with the Kalman filter, one updates deviations from the ensemble mean using a modified Kalman gain that is reduced in magnitude relative to the traditional Kalman gain. Deviations from the mean are thus reduced less in the analysis using $\widetilde{\mathbf{K}}$ than they would be using $\mathbf{K}$. In the canonical ensemble Kalman filter (EnKF; Burgers et al. 1998), the excess variance reduction caused by using $\mathbf{K}$ to update deviations from the mean is compensated for by the introduction of noise to the observations. In the EnSRF, the mean and departures from the mean are updated independently according to Eqs. (1) and (2). If observations are processed one at a time, the EnSRF requires about the same computation as the traditional EnKF with perturbed observations. The relative characteristics of the stochastic EnKF update algorithm and the deterministic EnSRF algorithm are discussed from different viewpoints in Whitaker and Hamill (2002) and Lawson and Hansen (2004).

Experience has shown that if the model state has more degrees of freedom than the ensemble has members, the background-error covariances cannot be directly estimated from a relatively small ensemble without producing a poor quality ensemble subject to the problem of filter divergence (Maybeck 1979 p. 338)

Hence, the background-error covariance model used in the data assimilation is typically modified by a process called "covariance localization" (e.g., Houtekamer and Mitchell 2001, Hamill et al. 2001). Mathematically, to apply covariance localization, the Kalman gain in (3) is replaced by a modified gain

$$\mathbf{K} = \left(\rho_S \circ \mathbf{P}^{\mathrm{b}}(t)\right)\mathbf{H}^{\mathrm{T}} \left(\mathbf{H}\left[\rho_S \circ \mathbf{P}^{\mathrm{b}}(t)\right]\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}, \tag{5}$$

where the operation $\rho_S \circ$ in (5) denotes a Schur product (an element-by-element multiplication) of a correlation matrix **S** with local support with the covariance model generated by the ensemble. The Schur product of matrices **A** and **B** is a matrix **C** of the same dimension, where $c_{ij} = a_{ij}\, b_{ij}$. For horizontal localization, one such correlation matrix can be constructed using an approximately Gaussian-shaped function described in Gaspari and Cohn (1999). When covariance localization is applied to smaller ensembles, it can actually result in more accurate analyses than would be obtained from larger ensembles without localization (Houtekamer and Mitchell 2001). Covariance localization, however, can introduce balance and consequent error-growth problems when the localization function is too narrow (Mitchell et al. 2002, Lorenc 2003).

b. Generating background forecasts

The other necessary part of the data assimilation cycle is the propagation of the forecast ensemble forward in time with the full, nonlinear forecast model to the time of the next new observations. This ensemble is used to estimate the mean state and background-error covariances. Unfortunately, the forecast model is not perfect, so even if one of the ensemble members happened to have a perfect initial condition, its subsequent forecast would contain errors. In Kalman filters (e.g., Gelb 1974, Maybeck 1979) the background-error covariances are specified by propagating the analysis-error covariances forward using the linear tangent **M** and its adjoint of the fully nonlinear forecast model operator $\mathcal{M}$, with an addition of covariance **Q** to account for model error:

$$\mathbf{P}^{\mathrm{b}}(t+1) = \mathbf{M}\mathbf{P}^{\mathrm{a}}(t)\mathbf{M}^{\mathrm{T}} + \mathbf{Q}. \tag{6}$$

$\mathbf{Q}$ is assumed to Gaussian, comprised of errors $\eta \sim (0, \mathbf{Q})$ that are uncorrelated in time and uncorrelated with internal error.

The EnKF and EnSRF achieve similar covariances to the Kalman filter, using the ensemble of fully nonlinear forecasts to estimate the background-error covariances. Let $\mathbf{X}^{\mathrm{a}}(t) = [\mathbf{x}_1^{\mathrm{a}}(t), \ldots, \mathbf{x}_n^{\mathrm{a}}(t)]$ be an ensemble of analyses at time t. Let $\mathcal{M}$ denote the foreward model operator between time t and $t+1$, and let $\overline{\mathcal{M}\mathbf{X}^{\mathrm{a}}}(t)$ denote an array composed of column vectors of ensemble mean forecasts started from time t, i.e., $\overline{\mathcal{M}\mathbf{X}^{\mathrm{a}}}(t) = [\overline{\mathcal{M}\mathbf{x}^{\mathrm{a}}}(t), \ \ldots \ , \overline{\mathcal{M}\mathbf{x}^{\mathrm{a}}}(t)]$, where $\overline{\mathcal{M}\mathbf{x}^{\mathrm{a}}}(t)) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{M}\mathbf{x}_i^{\mathrm{a}}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^{\mathrm{b}}(t+1)$. Then if forecast errors evolve linearly and the size n of the ensemble increases, $[\mathcal{M}\mathbf{X}^{\mathrm{a}}(t) - \overline{\mathcal{M}\mathbf{X}^{\mathrm{a}}}(t)]\ [\mathcal{M}\mathbf{X}^{\mathrm{a}}(t) - \overline{\mathcal{M}\mathbf{X}^{\mathrm{a}}}(t)]^{\mathrm{T}} = [\mathbf{X}^{\mathrm{b}}(t+1) - \overline{\mathbf{X}^{\mathrm{b}}}(t+1)]\ [\mathbf{X}^{\mathrm{b}}(t+1) - \overline{\mathbf{X}^{\mathrm{b}}}(t+1)]^{\mathrm{T}} \rightarrow \mathbf{M}\mathbf{P}^{\mathrm{a}}(t)\mathbf{M}^{\mathrm{T}}$. This indicates that the ensemble can be propagated forward and used to estimate the background-error covariances, minus the missing $\mathbf{Q}$ term in (6). If the forward propagator introduces systematic errors, the background forecasts should be adjusted; see Dee and Todling (2000). In this experiment, the systematic error was negligible. The subsequent experiments will demonstrate the impact of parameterizing $\mathbf{Q}$ through two general techniques, covariance inflation and/or additive errors.

With *covariance inflation*, ensemble members' deviations about their mean are inflated by an amount r, slightly greater than 1.0, before the first observation is assimilated:

$$\mathbf{x}_i^{\mathrm{b}}(t+1) \leftarrow r\Big(\mathbf{x}_i^{\mathrm{b}}(t+1) - \overline{\mathbf{x}}^{\mathrm{b}}(t+1)\Big) + \overline{\mathbf{x}}^{\mathrm{b}}(t+1). \tag{7}$$

Here, the operation $\leftarrow$ denotes a replacement of the previous value of $\mathbf{x}_i^{\mathrm{b}}(t+1)$. Application of a moderate inflation factor has been found to improve the accuracy of assimilations in perfect-model experiments (Hamill et al. 2001, Whitaker and Hamill 2002) and real-data simulations (W04). Note that inflation increases the spread of the ensemble, but it does not change the subspace spanned by the ensemble. The common assumption in the Kalman filter derivation is that model error and internal error growth are uncorrelated. If indeed the model error projects into a substantially different subspace than the ensemble, this parameterization may not be effective. However, if the dynamically relevant part of model error is part that projects onto the growing

modes spanned by the ensemble, covariance inflation may be effective. Another issue with covariance inflation is that a single inflation factor r may not be optimal over all parts of a model domain. Consider, for example, a situation where observations are very plentiful in one hemisphere and nearly nonexistant in the other. In the data-sparse hemisphere, observations will not reduce the spread of the ensemble as much during the updates, and covariance inflation may lead to an unbounded growth of ensemble variance.

Additive errors may avoid some of these problems. Noise η_i with the same dimension as the model state is added to each ensemble member background forecast before the update cycle starts:

$$\mathbf{x}_i^{\mathrm{b}}(t+1) \leftarrow \mathbf{x}_i^{\mathrm{b}}(t+1) + \eta_i. \tag{8}$$

η_i ought to sample the probability distribution of accumulated model errors $\mathbf{Q}$, i.e., $\langle \eta_i \eta_i^{\mathrm{T}} \rangle = \mathbf{Q}$. Also, $\langle\ \mathbf{x}'^{\mathrm{b}}_i(t+1)\ \eta_i^{\mathrm{T}} \rangle = 0$; that is, the model error ought to be independent of the internal error, though the model error might still be dependent on the model state. Unlike covariance inflation, the resulting modified ensemble may well span a somewhat different subspace than the unmodified ensemble, and additive error will not have the tendency to excessively inflate the forecast variance when no observations are present, for the additive errors have a proportionally larger influence where background-error variances are small compared to where they are large.

In the subsequent section we describe tests of ensemble assimilations with additive errors generated using three different methods. The first approach was to sample differences between the resolved scales of model forecasts at different resolutions. Recall that in our experiment, the true model state was from a T127 simulation and the ensemble forecasts conducted at T31. Suppose generating a T127 simulation is a computational impossibility, but we have the ability to generate a long T63 simulation. The difference of the resolved scales between T63 and T31 forecasts can then be computed: $\mathcal{M}^{31}(\mathcal{T}[\mathbf{x}^{63}(t)]) - \mathcal{T}[\mathbf{x}^{63}(t+1)]$. This time series can be randomly sampled, and a different sample η_i can be added to the each ensemble member according to (8). In fact, what typically was added was $s\ \eta_i$, where s represented a scaling factor greater

than 1.0. That is, the additive error samples were inflated in size somewhat before being added to the background forecasts.

Another ad-hoc approach that was tested was to add scaled-down differences between random model states from the forecast model climatology, i.e., $\eta_i = s\,[\mathbf{x}^{31}(t_r) - \overline{\mathbf{x}}^{31}]$, where t_r is a random time from the time series of the T31 nature run, $\overline{\mathbf{x}}^{31}$ is the mean climatological state, and again s is a scaling factor. A similar approach was used for generating ensemble perturbations by Schubert and Suarez (1989). A third additive error approach used scaled-down short-term (24-h) forecast tendencies: $\eta_i = s\,[\mathbf{x}^{31}(t_r) - \mathbf{x}^{31}(t_r - 24h)]$, where again t_r is a random time from the time series.

Figure 5 shows the spatially lagged correlations of additive errors generated through these three processes, as well as the lagged correlation of the true model errors, i.e., $\mathcal{M}^{31}(\mathcal{T}[\mathbf{x}^{127}(t)]) - \mathcal{T}[\mathbf{x}^{127}(t+1)]$. The differences between model forecasts at T63 and T31 resolutions produced additive errors with the shortest correlation length scales, very similar to the correlation length scale of the actual model error. The length scales from 24-h forecast tendencies were substantially longer, and samples from climatology had the longest length scales. By generating additive error using these three methods, we can examine whether overestimating the correlation length scale of additive errors will reduce the accuracy of the subsequent analyses. Figure 5 also shows the correlation structure of typical background forecasts from one of the subsequent ensemble data assimilation experiments (experiment 5, described in Section 4b). Additive errors from differences between T63 and T31 forecasts shortened the correlation length scale of the background errors, while climatological additive noise lengthened it.

4. EXPERIMENT DESIGN

a. Observations, ensemble configuration, and evaluation technique

We now describe a large number of data assimilation experiments. In each experiment, observations of $\Delta\pi_2$ were taken at a set of nearly equally spaced locations on a spherical geodesic grid (Fig. 6). The observations consisted of the T127 true state plus errors drawn from a dis-

tribution with zero mean and standard deviation of 8.75 $Jkg^{-1}K^{-1}$. Observation errors were independent at each location and time, and observations were assimilated every 24 h.

In all the experiments, the ensemble size was 208 members. The ensemble was initialized with random draws from the climatology of the assimilating model. The data assimilation proceeded over a 150-day period; the first 50 days were discarded as a spinup period, and error statistics were accumulated over the remaining 100 days. Aside from experiment 1, a perfect-model simulation at T127, all the other experiments were conducted at T31 resolution.

After some preliminary experimentation, the chosen covariance localization for all experiments used the Gaspari and Cohn (1999) horizontal localization function that decayed to zero at 5000 km. This length scale was chosen after trying several cutoff distances in perfect-model experiments with the T127 model (experiment 1 below).

The ensemble-mean analysis error and spread (standard deviation of the ensemble about its mean) were measured in three norms, a globally averaged, mass-weighted kinetic energy norm, an upper-layer thickness norm (since the model top is fixed, this is equivalent to a measurement of interface Exner function), and surface Exner function norm. The kinetic energy norm at a given time is

$$\| \cdot \|_{ke} = \left[\frac{\int_S \int_L \pi \left(\mathbf{u}'^2 + \mathbf{v}'^2 \right) \, dS dL}{\int_S \int_L \pi \, dS dL} \right]^{0.5} \tag{9}$$

Here, $\mathbf{u}'$ and $\mathbf{v}'$ denote the model wind components' ensemble-mean error or spread, π again refers to the model state's Exner function, S refers to the integration over the sphere, and L integration over the model layers. The interface height norm is

$$\| \cdot \|_{\Delta\pi_2} = \frac{1}{A} \left[\int_S \Delta\pi_2'^2 \, dS \right]^{0.5} \tag{10}$$

where $\Delta\pi_2'$ denotes the upper-layer Exner function thickness ensemble-mean error or deviation from the mean, and A is the area covering the earth. The surface Exner function norm is

$$\| \cdot \|_{\Delta\pi_{1+2}} = \frac{1}{A} \left[\int_S \left(\Delta\pi_1' + \Delta\pi_2' \right)^2 \, dS \right]^{0.5} \tag{11}$$

where $\Delta\pi'_1$ is the lower-layer Exner function thickness ensemble-mean error or perturbation.

Model error was a very prominent component of the errors in these experiments, arguably more than they are in operational ensemble forecasting (e.g., Simmons and Hollingsworth 2002, their Fig. 6). Figure 7 examines the growth of ensemble forecast spread compared to the ensemble mean error, taken from Experiment 4, described below. Compared with the figure from Simmons and Hollingsworth, it is apparent that the growth of spread lags far behind the ensemble mean error, significantly more than it does in a recent version of ECMWF's model.

b. Experiments

What now follows are a description of each data assimilation experiment, summarized in Table 1.

Experiment 1, "*T127 perfect model*," performed the ensemble data assimilations using the EnSRF in a perfect-model context, assimilating the T127 observations using an ensemble of forecasts at T127 with $r = 1.01$ (some covariance inflation has been found to be helpful in perfect-model simulations; see Hamill et al. 2001, Whitaker and Hamill 2002). To be consistent with the rest of the data assimilation experiments conducted at T31 resolution, the error and spread statistics were evaluated only for scales T31 and larger.

Experiment 2, "*covariance inflation*," used an EnSRF algorithm with model error parameterized with a straightforward covariance inflation. Various values of covariance inflation were tried, but all simulations had the tendency over many weeks to become numerically unstable. For subsequent discussion, assume the covariance inflation is $r = 1.08$, i.e., perturbations were inflated by 8 % at the beginning of each assimilation cycle.

Experiment 3, "*restarted covariance inflation*," was a modification of experiment 2. Since covariance inflation algorithms typically outperformed 3D-Var experiments (described later) for a short period of time, we considered an experiment where the EnSRF was re-run in overlapping 20-day windows. The first 10 days in each window were discarded as a spinup period, and statistics accumulated for the remaining 10 days. Thus, the overall 100 days where statistics were accumulated represented 10 overlapping data assimilation experiments. In this experiment, $r = 1.10$.

Experiment 4, "*T127 additive error,*", used additive errors and no covariance inflation. In this case, additive errors for each member were randomly drawn from a time series of the true additive errors, found by comparing the T127 and T31 simulations. This simulation was used as a basis for comparing how skillful other additive error parameterizations were.

Experiment 5, "*T63 additive*," used only additive errors randomly drawn from differences between T63 and T31 simulations. The scaling factor s for inflating additive model error samples was set to 1.20, chosen by experimentation.

Experiment 6, "*T63 analogs*," set $s = 1.20$ as in experiment 5, but in this experiment non-random times were selected from the time series of differences between T63 and T31 simulations. At each data assimilation time, we searched a 10,000 day time series of initial conditions and found the 208 times in that time series that were the closest fit to the observations at this time. The analogs were a somewhat better fit to the observations, typically around 20-40% closer fit than the random model states. Since model error appeared to be highly state dependent (Fig. 3), the hypothesis was that selecting additive model error samples from relatively similar states would provide a better parameterization than from random states.

Experiment 7, "*24-h tendency*:" additive errors were randomly drawn from a sample of differences between model states separated by 24 h. These additive errors were scaled by 0.25.

Experiment 8, "*climatology*:" additive errors were randomly drawn from a sample of anomalies from the model's climatological mean state, scaled by 0.25.

Experiment 9, "*3D-Var*," was an experiment simulating a 3-dimensional variational analysis by updating a single model state using a static background-error covariance estimate. The static background-error covariances were formed from an ensemble consisting of 208 random samples of actual background error from experiment 5 inflated by 40 %. The same 5000 km localization of covariances was used. A similar approach to simulating 3D-Var in an ensemble filter framework was described in Zhang and Anderson (2003) and Evensen (2003).

5. RESULTS

Table 2 summarizes the time-averaged ensemble-mean error and spread of the analyses. For reference, the errors of climatology were approximately 28 ms^{-1} in the kinetic-energy norm. The accuracy of the resolved scales from Experiment 1, the T127 Perfect Model, produced the lowest error in all norms. Experiment 2, a straightforward application of covariance inflation, became numerically unstable. However, if the ensemble was regularly re-started, as it was in Experiment 3, then the assimilations produced analyses with $\sim$ 7, 15, and -18 % relative improvements, measured as the fractional difference in error between 3D-Var (Exp. 10) and the T127 Perfect Model (Exp. 1) using the three norms in eqs. (9)-(11).

The T127 additive error, the T63 additive and T63 analogs experiments all had nearly the same errors. These produced $\sim$ 41, 38, and 85 % relative reductions in error for the T63 additive experiment. Interestingly, the parameterization using the T63 runs were equally skillful to using T127 samples of the additive error (Exp. 4), indicating that the use of model differences from T63 simulations was a highly effective choice of parameterization. The use of closer analogs of model error (Exp. 6) did not have a noticeable impact on model accuracy.

Why was the covariance inflation experiment substantially less accurate than these additive error experiments? Consider a snapshot 10 days into the verification from the T63 additive error experiment (Fig. 8). This plot shows at each grid point the Kalman gain (eq. 3) for correcting (a) the upper-layer thickness $\Delta\pi_2$ and (b) the lower-layer thickness $\Delta\pi_2$ based upon that same grid point's upper-layer thickness $\Delta\pi_2$. This plot indicates that in the tropics, the assimilation of observations of upper-layer thickness typically resulted in small corrections to the lower-layer thickness, which is largely anticorrelated with the correction to upper-layer thickness. Hence, the lower-layer thickness ensemble variance was not substantially decreased during the update in the tropics. However, during the subsequent forecast, the ensemble spread grew in the extratropics due to internal error growth and was further inflated in the extratropics by the additive model error parameterization (e.g., samples such as in Fig. 3). Figure 9(a) shows that the zonally averaged background spread in the T63 additive simulation reaches a quasi-equilibrium state, decreased in the update but growing by a similar amount during the forecast.

If we consider the covariance inflation simulation, we see a runaway growth of spread in the tropics (Fig. 9b). Covariance inflation caused the spread prior to the update to increase proportionally everywhere. Repeated inflation every analysis cycle eventually produces an ensemble with unrealistically large spread in the tropics. Since analysis increments are proportional to the background spread at the observation location, this implies that tropical observations then began to introduce large adjustments to the background. Increasingly with time, these increments became less realistic. In this particular model, for instance, the unrealistic increments can be seen by examining surface Exner function, π_{top} + $\Delta\pi_1$ + $\Delta\pi_2$. For our additive error simulation, the increments to layer thicknesses $\Delta\pi_1$ and $\Delta\pi_2$ were strongly anticorrelated (Fig. 8), limiting the size of changes to column thickness and hence surface Exner function. In the covariance inflation sumulation, the ensemble lost this anticorrelation, and unrealistically large adjustments to total column thickness resulted (Fig. 10b). This behavior of the ensemble filter was somewhat different than the classical "filter divergence" problem. Here, the problem was not that the background-error covariances are systematically underestimated so that the observations are effectively ignored. Rather, errors increased because the background-error covariance model became corrupted due to the effects of covariance inflation in regions where observations have little or no influence. These results show that a uniform, large covariance inflation factor to parameterize model error is likely to cause data assimilation problems if the observations themselves are not uniform in density. This situation can be expected when background errors differ in magnitude and growth rate from one part of the domain to another; see also the discussion in Snyder and Zhang (2003, sec. 6b). A location-dependant inflation factor, akin to the "mask" used in the breeding technique of Toth and Kalnay (1997), could in principle ameliorate some of these problems, or the approach of Zhang et al. (2004; see eq. 5).

We next consider whether forecast accuracy varied with the specific type of additive model error parameterization. Parameterizing the model error with 24-h tendencies and climatology did produce progressively less accurate assimilations. 24-h tendencies produced analyses that were 33, 34, and 50 % improvements in relative errors, while additive errors sampled from climatol-

ogy produced assimilations that were 10 and 31 % better and 44 % worse in the three norms (recall that the T63 additive experiment produced 41, 38, and 85 % reductions). This result indicates that the spatial structure of additive errors were determinative of the accuracy of ensemble data assimilations. Here, the source of model error was model truncation, introducing errors that were relatively small in scale. When the additive errors samples were inappropriately large in scale, they modified the background ensemble so its perturbations were also too large in scale. When assimilating a single observation, the analysis increment is proportional to the background error covariance between a grid point and the observation location. Consequently, an overly large spatial scale of background-error covariances caused the observations to have unduly large influence in adjusting the analysis at distances far from the observation location, and the skill of the assimilations decreased.

A characteristic seen in nearly all the ensemble simulations in Table 2 is that the spread for the surface Exner function is much less than the ensemble-mean error, even in the perfect-model experiment. The surface Exner function is not actually a state variable, but is diagnosed from the sum of layer thicknesses. We suspect that the larger deficiency in spread is due to larger errors in these diagnosed cross covariances of the state variables, for the spread deficiency was noted even in the T127 perfect model simulation.

6. DISCUSSION AND CONCLUSIONS

In these experiments, we considered different methods for parameterizing the model error in ensemble assimilations due to truncation of the model resolution. Of course, model errors in full numerical weather prediction models can caused by many other factors, such as improper parameterizations of other sub-gridscale processes such as convection not included in this simplified model. Our choice here was to try to understand one relatively simple source of error and explore which methods were most effective for treating them. The experimental setup was designed to produce large model errors, probably a larger contribution to forecast error than occurs in current-generation numerical models.

The experiment was conducted with a 2-layer primitive equation model. The true state was a T127 forecast nature run. Ensemble data assimilations were performed with the same model at T31 resolution, assimilating imperfect observations drawn from the T127 forecast. Several methods were considered for parameterizing the model error at the resolved scales (T31 and larger) due to interaction with the unresolved scales (T32 to T127). "Covariance inflation" simply inflated the background forecast members' deviations about the ensemble mean. Another, "additive errors," added specially structured noise to each ensemble member.

The method of parameterizing this model error had a substantial effect on the accuracy of ensemble data assimilations. The best additive model error parameterization was able to reduce the relative error between 3D-Var and a near-perfect assimilation system by 38 to 85 %, depending on the norm used to measure the error. Covariance inflation produced ensembles with analysis errors that were typically slightly less than the analysis errors from 3-dimensional variational (3D-Var) assimilation, but in order for the method to remain stable, the assimilations had to be continually restarted. An examination of the characteristics of the covariance inflation run showed that it developed perturbations that produced an inaccurate model of the background error covariances.

The most effective additive error parameterization used samples of model error from a time series of differences between T63 and T31 forecasts. Scaled samples of differences between model forecast states separated by 24 h were also tried, as well as scaled samples of the model state's anomaly from the model climatology. These latter two methods of generating additive error samples produced analyses that were less accurate than when using differences between T63 and T31 forecasts but more accurate than covariance inflation. Differences between T63 and T31 forecasts produced an additive error parameterization with a length scale very similar to the true model error length scale, while the scaled 24-h forecast and deviations from climatology had progressively longer length scales, indicating that the correlation length scale of the additive errors is important.

This method of parameterizing model error does not use innovation statistics from the data assimilation, as proposed by Dee (1995), Dee and da Silva (1998, 1999) and tested in an ensemble filter by Mitchell and Houtekamer (2000). Use of innovation statistics has the concrete benefit in that model errors are estimated using independent information, the observations. The model error parameters may also be adaptively tuned, increasing the magnitude of model error when the first guess is not a good fit to the observations. A drawback of the method is that it may not be possible to estimate a large number of parameters of a model-error covariance model using this approach, and the method depends on the observing network and an accurate characterization of the observation errors. Indeed, though Mitchell and Houtekamer's approach was demonstrated to work well in an observing system simulation experiment context, in practice they now generate additive errors to be consistent with random draws of 3D-Var covariances (Houtekamer et al. 2004).

Unlike the innovation statistics method, here the differences between short-term model integrations were used to generate samples of model error. As such, the additive error samples represented actual realizations of model differences, with structures that had consistent mass/wind relationships and appropriate spatial structure. In these experiments we knew a priori what the model error was and thus could design a method of using differences between forecast models to accurately simulate these model errors. In practice, the actual model deficiencies will not be very well known. They are likely to be an amalgamation of errors from many sources: errors in convective parameterizations, boundary-layer parameterizations, radiation, cloud microphysics, land-surface processes, parameter mis-estimation, model truncation, and so on. However, our approach is appealing at least for its simplicity; if one has a well-founded reason for believing that a particular parameterization is problematic, two models with different parameterizations could be run and the resulting differences used as samples of additive errors. Further, if additive errors come from many sources, there is no conceptual reason why multiple, independent additive errors could not be added to each ensemble member, one for each suspected type of model error. Perhaps innovation statistics approach could be could be incorporated as well,

adaptively changing the amount of model error added depending on the misfit of the first guess to the observations. Clearly, methods for parameterizing model error are important and deserving of much more exploration.

7. ACKNOWLEDGMENTS

The lead author's participation was partially supported under National Science Foundation (NSF) grants ATM-0130154, ATM-0112715, and ATM-0205612. Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily represent the views of the NSF. The NOAA/FSL supercomputer resource hosted this experiment. Chris Snyder (NCAR) provided a helpful informal review of this manuscript, and Jeff Anderson (NCAR) is thanked for his consultations on ensemble assimilation issues.

REFERENCES

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884-2903.

——— , 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634-642.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719-1724.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367-1387.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128-1145.

——— , and A. M. da Silva, 1998: Data assimilation in the presence of forecast bias. *Quart. J. Roy. Meteor. Soc.*, **124**, 269-295.

——— , and ——— , 1999: Maximum-likelihood estimation of forecast parameters for atmospheric data assimilation. Part I: methodology. *Mon. Wea. Rev.*, **127**, 1822-1834.

——— , and R. Todling, 2000: Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Mon. Wea. Rev.*, **128**, 3268-3282.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10143-10162.

——— , and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85-96.

——— , 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343-367.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723-757.

Gelb, A. (ed.), 1974: *Applied optimal estimation.* MIT Press, 374 pp.

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter / 3d-variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

——— , J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776-2790.

——— , and C. Snyder, 2002: Using improved background error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, **130**, 1552-1572.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

——— , ——— , 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

——— , ——— , G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2004: Atmospheric data assimilation with the ensemble Kalman filter: results with real observations. *Mon. Wea. Rev.*, in press.

Lawson, W. G., J. A. Hansen, 2004: Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.*, in press.

Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, , **38A**, 97-110.

Legras, B., and R. Vautard, 1996: A guide to Lyapunov vectors. *Proceedings of the ECMWF Seminar on Predictability*, 4-8 September 1995, vol 1, 143-156. Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, United Kingdom].

Leith, C. E., and R. H. Kraichnan, 1972: Predictability of turbulent flows. *J. Atmos. Sci.*, **29**, 1041-1058.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183-3204.

Lorenz, E. N., 1969: the predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289-307.

Maybeck, P. S., 1979: *Stochastic models, estimation, and control. Volume 1*. Academic Press, 423 pp.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416-433.

——— , and ——— , and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 2791-2808.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: a proposal for non-local stochastic-dynamic parameterization in weather and climate prediction models. *Quart. J. Roy. Meteor. Soc.*, **127**, 279-304.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747-1763.

Potter, J., 1964: W matrix augmentation. M.I.T. Instrumentation Laboratory Memo SGA 5-64, Massachusetts Institute of Technology, Cambridge, MA.

Rabier, F., J.-N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1-39.

——— , H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143-1170.

Schubert, S. D., and M. Suarez, 1989: Dynamical predictability in a simple general circulation model: average error growth. *J. Atmos. Sci.*, **46**, 353-370.

Simmons, A. J., and A. Hollingsworth, 2002: Some aspects of the improvement in skill of numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **128**, 647-677.

Snyder, C., and T. M. Hamill, 2003: Leading Lyapunov vectors of a turbulent baroclinic jet in a quasigeostrophic model. *J. Atmos. Sci.*, **60**, 683-688.

——— , and F. Zhang, 2003: Assimilation of simulated doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663-1677.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131** , 1485-1490.

Toth, Z, and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Tribbia, J. J., and D. P. Baumhefner, 2004: Scale interactions and atmospheric predictability: an updated perspective. *Mon. Wea. Rev.*, **132**, 703-713.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, , **130**, 1913-1924.

——— , G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data asssimilation. *Mon. Wea. Rev.*, in press.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238-1253.

Zhang, S., and J. L. Anderson, 2003: Impact of spatially and temporally varying estimates of error covariance on assimilation in a simple atmospheric model. *Tellus*, **55A**, 126-147.

Zou, X., A. Barcilon, I. M. Navon, J. Whitaker, and D. G. Cacuci, 1993: An adjoint sensitivity study of blocking in a two-layer isentropic model. *Mon. Wea. Rev.*, **121**, 2833-2857.

FIGURE CAPTIONS

Figure 1. Zonal mean statistics for nature runs at T127, T63, and T31 resolutions. (a) Upper-layer zonal wind, (b) lower-layer zonal wind, and (c) Upper-layer $\Delta\pi$.

Figure 2. Kinetic energy spectrum of nature runs at T127, T63, and T31 resolutions. Curves for T63 and T31 are shifted by one and two orders of magnitude, respectively.

Figure 3. Time series of upper layer thickness 24-h model errors. True state of upper layer $\Delta\pi$ resolved scales overplotted in solid lines (contour interval 50 $J\ kg^{-1}\ K^{-1}$). Geographical features are overlaid only for perspective.

Figure 4. Kinetic energy spectrum of model errors due to lack of interactions with unresolved scales, accumulated at 3, 24, 48, 72, 96, and 120 h. (a) Spectrum of model errors when model is truncated at T31, and (b) spectrum when truncated at T63.

Figure 5. Spatially lagged correlation along 45^{o} N latitude of $\Delta\pi$ as a function of zonal distance for various additive error models, as well as the correlation of the background forecast ensemble.

Figure 6. Observation locations. Observations were generated globally on a spherical geodesic grid. There are 362 observation locations worldwide.

Figure 7. Growth of ensemble spread and ensemble-mean error in the kinetic-energy norm, taken from the perfect additive error experiment (#4).

Figure 8. Upper-layer ensemble mean background $\Delta\pi_2$ and the Kalman gain for (a) the upper layer $\Delta\pi_2$ and (b) the adjacent lower layer $\Delta\pi_1$ at time t=10 days from the T63 additive experiment.

Figure 9. Hovmoller diagrams of zonally averaged background spread of $\Delta\pi_1$ from the (a) T63 additive experiment, and (b) covariance inflation experiment.

Figure 10. Hovmoller diagrams of zonally averaged Kalman gain **K** of changes of surface Exner function $\Delta\pi_1 + \Delta\pi_2$ from the (a) T63 additive experiment, and (b) covariance inflation experiment.

Table 1. Table of data assimilation experiments performed. The first column denotes the experiment number, the second column the name of the experiment. The third column denotes the general data assimilation methodology. The fourth column indicates the amount of covariance inflation, r. The fifth column indicates the amount rescaling of additive error, s. The sixth column indicates the additive error type. The seventh column indicates whether additive errors in EnSRF assimilations were subjected to a bias correction.

Exp #	*Name*	*Method*	*r*	*s*	*Additive Error Type*	*Bias Correction?*
1	T127 perfect model	EnSRF	1.01	n/a	n/a	n/a
2	Covariance inflation	EnSRF	1.08	n/a	n/a	n/a
3	Restarted cov inflation	EnSRF	1.10	n/a	n/a	n/a
4	T127 additive error	EnSRF	n/a	1.00	T127	Yes
5	T63 additive	EnSRF	n/a	1.20	T63	Yes
6	T63 analogs	EnSRF	n/a	1.20	T63 global analogs	Yes
7	24-h tendency	EnSRF	n/a	0.25	24 h tendency	No
8	Climatology	EnSRF	n/a	0.25	Climatology	No
9	3D-Var	3D-Var	n/a	1.40	n/a	n/a

Table 2. Global ensemble-mean analysis error and ensemble spread, measured in kinetic energy norm, upper-layer Exner-function thickness norm, and surface Exner function norm.

Exp #	*Name*	Err_{ke}	Spr_{ke}	$Err_{\Delta\pi_2}$	$Spr_{\Delta\pi_2}$	$Err_{\Delta\pi_{1+2}}$	$Spr_{\Delta\pi_{1+2}}$
1	T127 perfect model	3.67	3.96	5.37	5.51	1.17	0.97
2	Covariance inflation	–	–	(diverged)	–	–	–
3	Restarted cov inflation	5.62	5.47	7.74	5.04	1.83	2.31
4	T127 additive error	4.92	4.85	7.10	6.81	1.25	0.86
5	T63 additive	4.93	4.81	7.14	6.62	1.28	0.82
6	T63 analogs	4.96	4.76	7.10	6.52	1.27	0.82
7	24-h tendency	5.08	4.18	7.21	5.53	1.45	1.14
8	Climatology	5.56	5.36	7.30	5.89	1.98	1.93
9	3D-Var	5.76	n/a	8.16	n/a	1.73	n/a

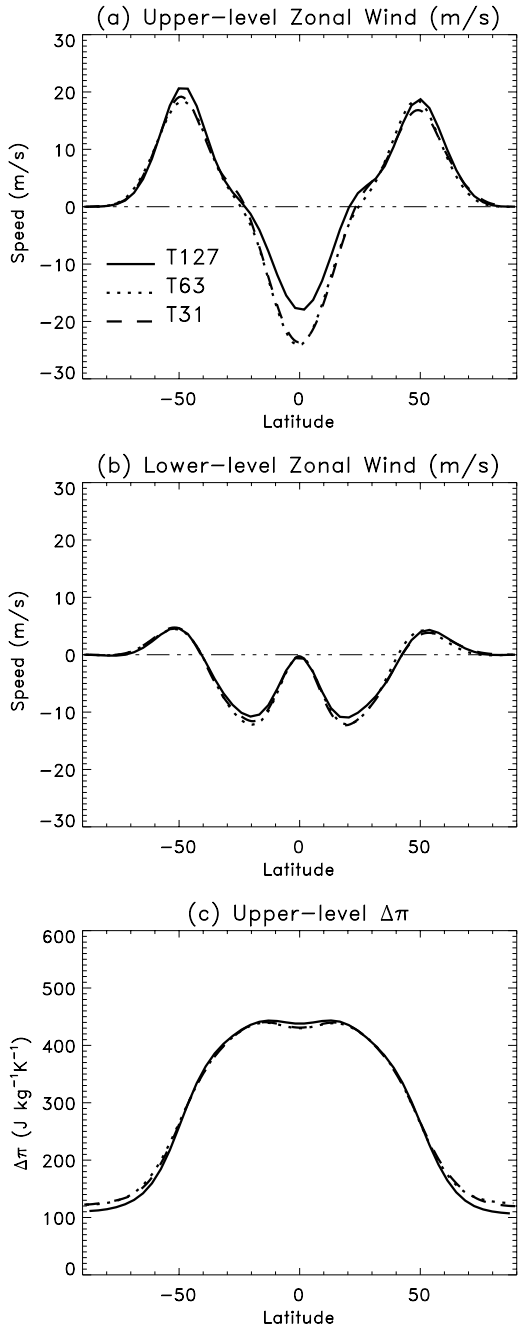

Figure 1. Zonal mean statistics for nature runs at T127, T63, and T31 resolutions. (a) Upper-layer zonal wind, (b) lower-layer zonal wind, and (c) Upper-layer $\Delta\pi$.

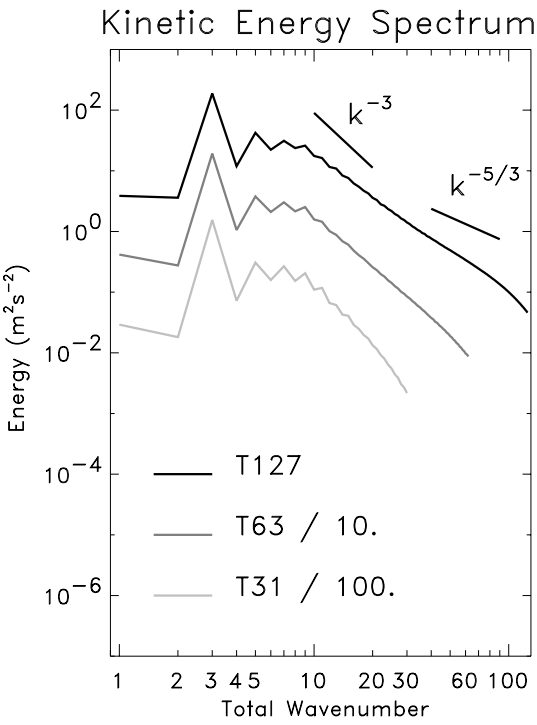

Figure 2. Kinetic energy spectrum of nature runs at T127, T63, and T31 resolutions. Curves for T63 and T31 are shifted by one and two orders of magnitude, respectively.

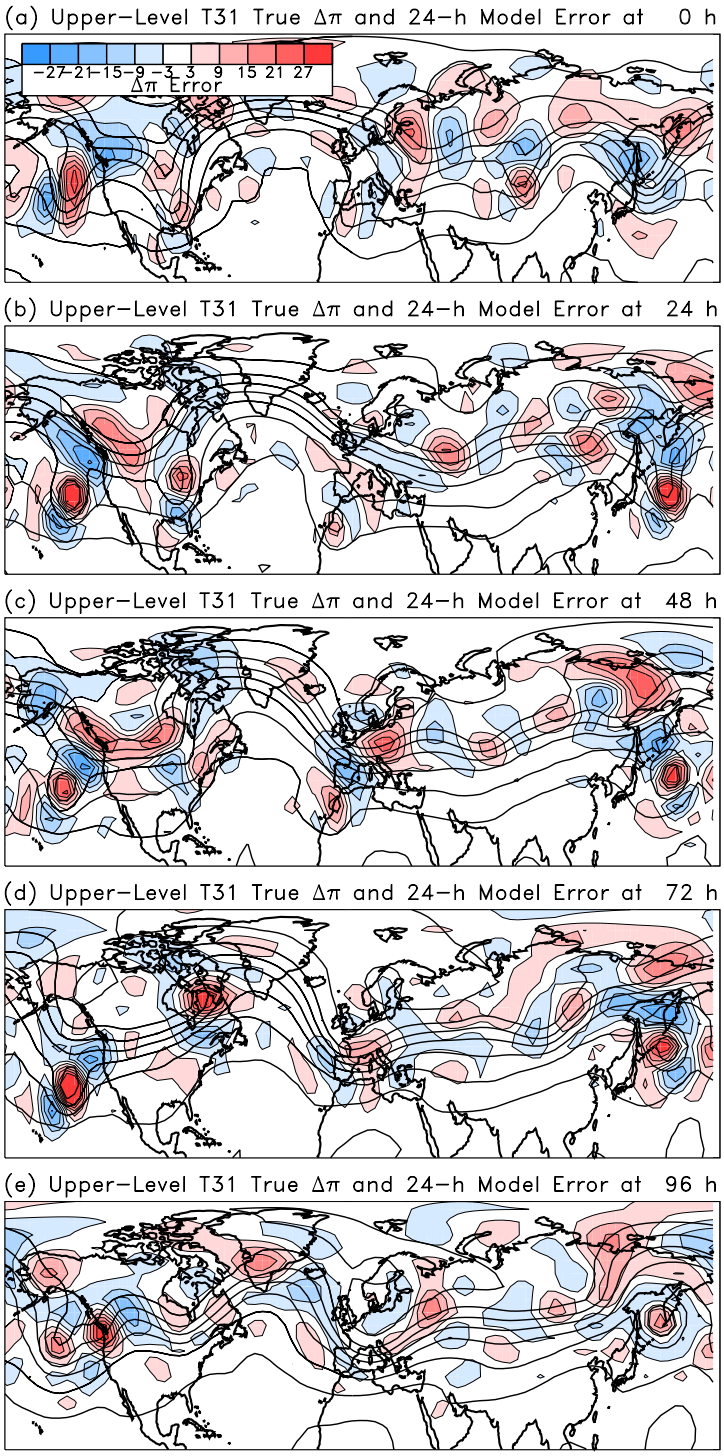

Figure 3. Time series of upper layer thickness 24-h model errors. True state of upper layer $\Delta\pi$ resolved scales overplotted in solid lines (contour interval 50 $J\ kg^{-1}\ K^{-1}$). Geographical features are overlaid only for perspective.

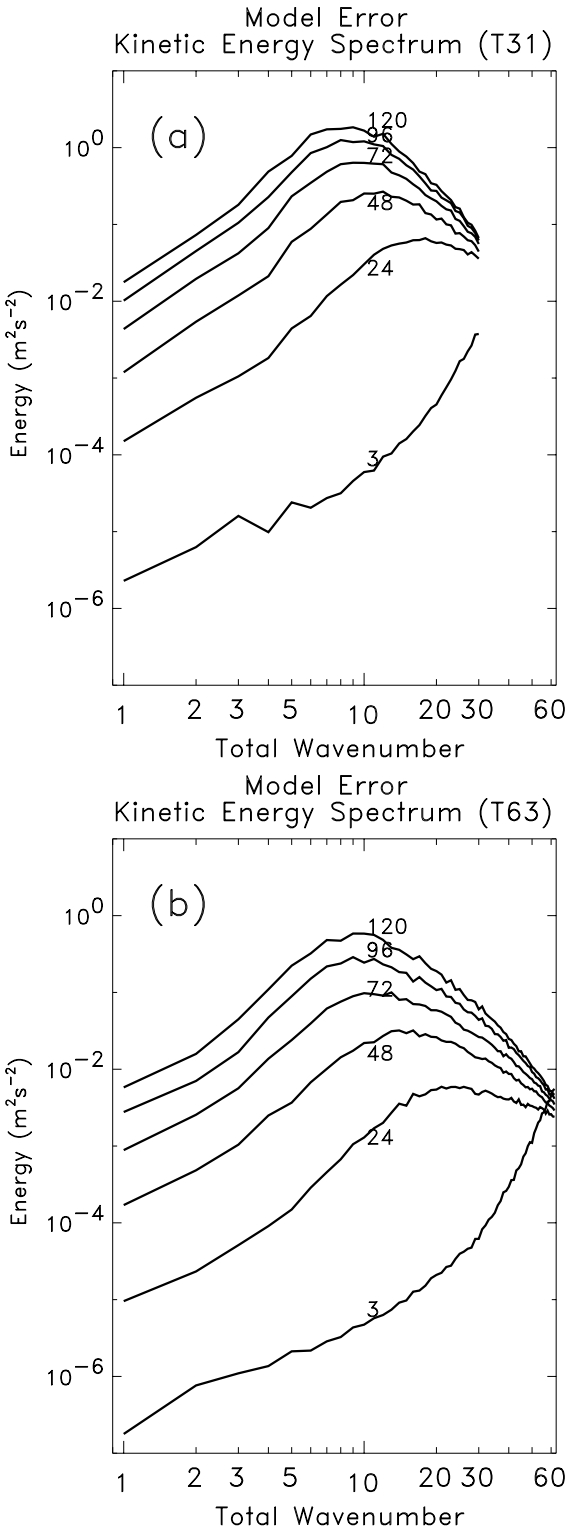

Figure 4. Kinetic energy spectrum of model errors due to lack of interactions with unresolved scales, accumulated at 3, 24, 48, 72, 96, and 120 h. (a) Spectrum of model errors when model is truncated at T31, and (b) spectrum when truncated at T63.

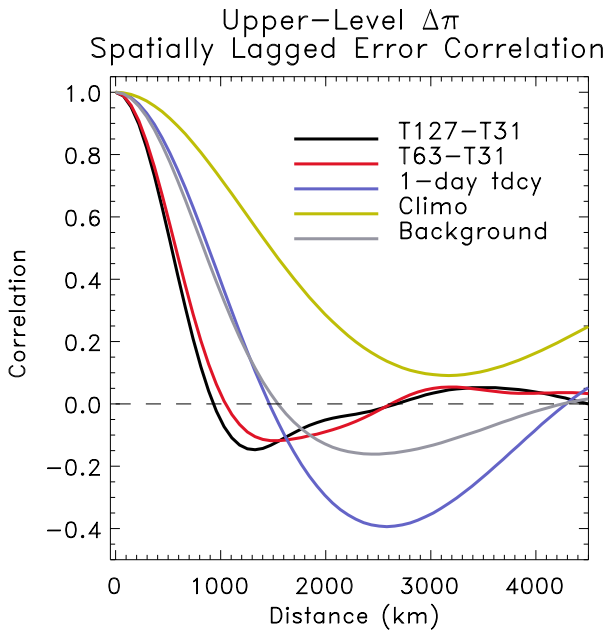

Figure 5. Spatially lagged correlation along 45^{o} N latitude of $\Delta\pi$ as a function of zonal distance for various additive error models, as well as the correlation of the background forecast ensemble.

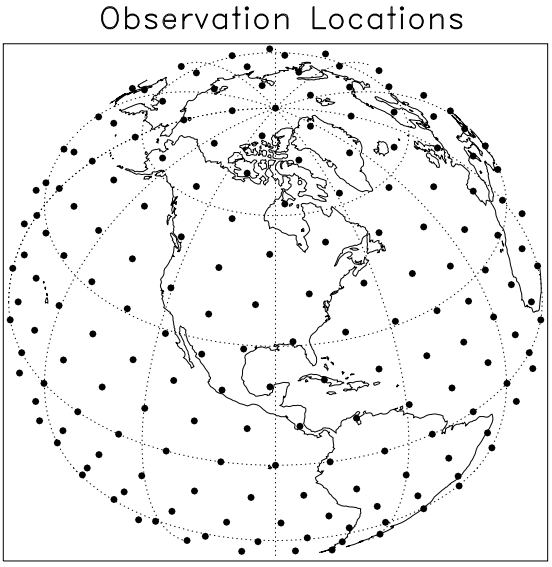

Figure 6. Observation locations. Observations were generated globally on a spherical geodesic grid. There are 362 observation locations worldwide.

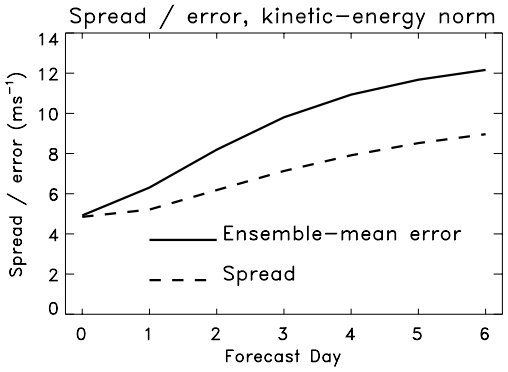

Figure 7. Growth of ensemble spread and ensemble-mean error in the kinetic-energy norm, taken from the perfect additive error experiment (#4).

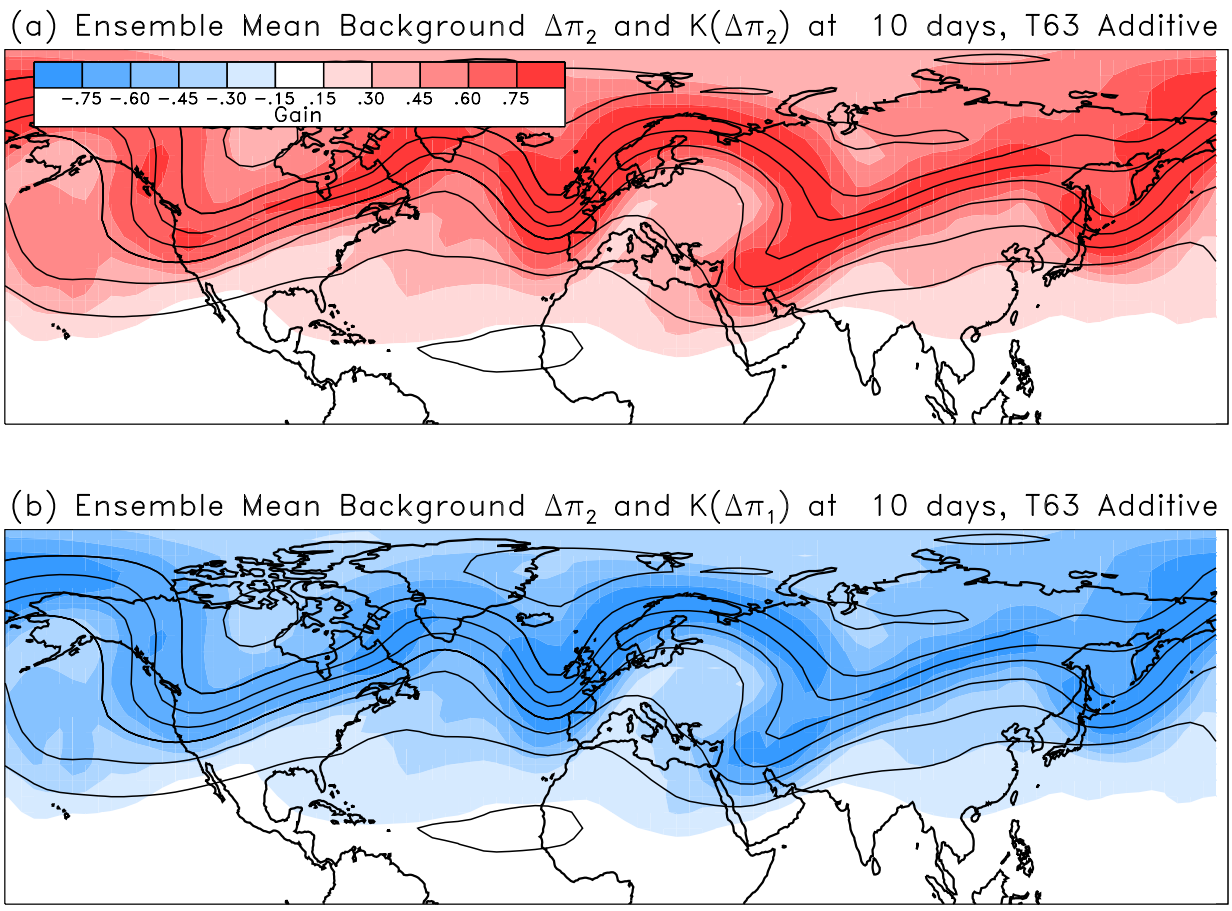

Figure 8. Upper-layer ensemble mean background $\Delta\pi_2$ and the Kalman gain for (a) the upper layer $\Delta\pi_2$ and (b) the adjacent lower layer $\Delta\pi_1$ at time t=10 days from the T63 additive experiment.

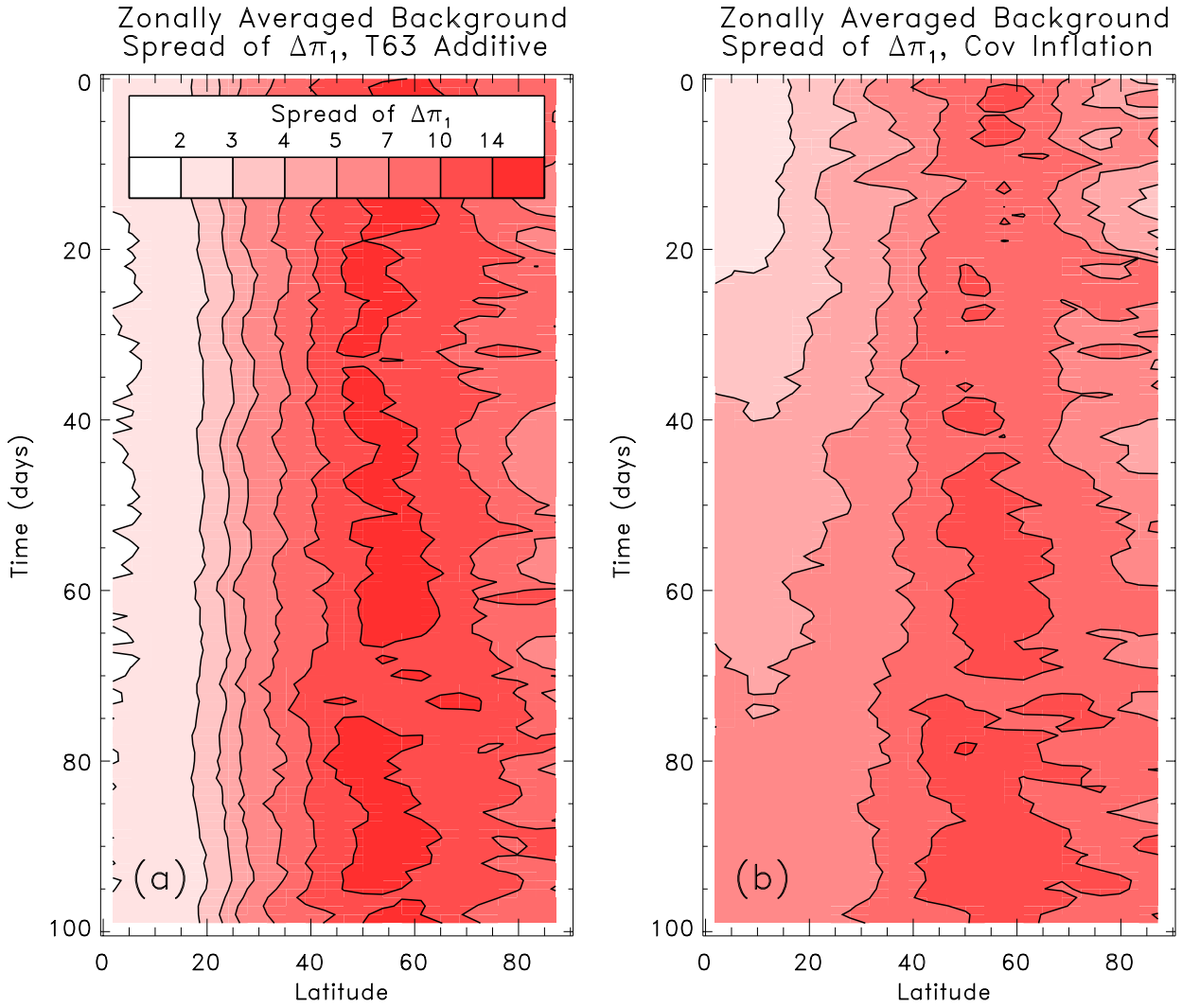

Figure 9. Hovmoller diagrams of zonally averaged background spread of $\Delta\pi_1$ from the (a) T63 additive experiment, and (b) covariance inflation experiment.

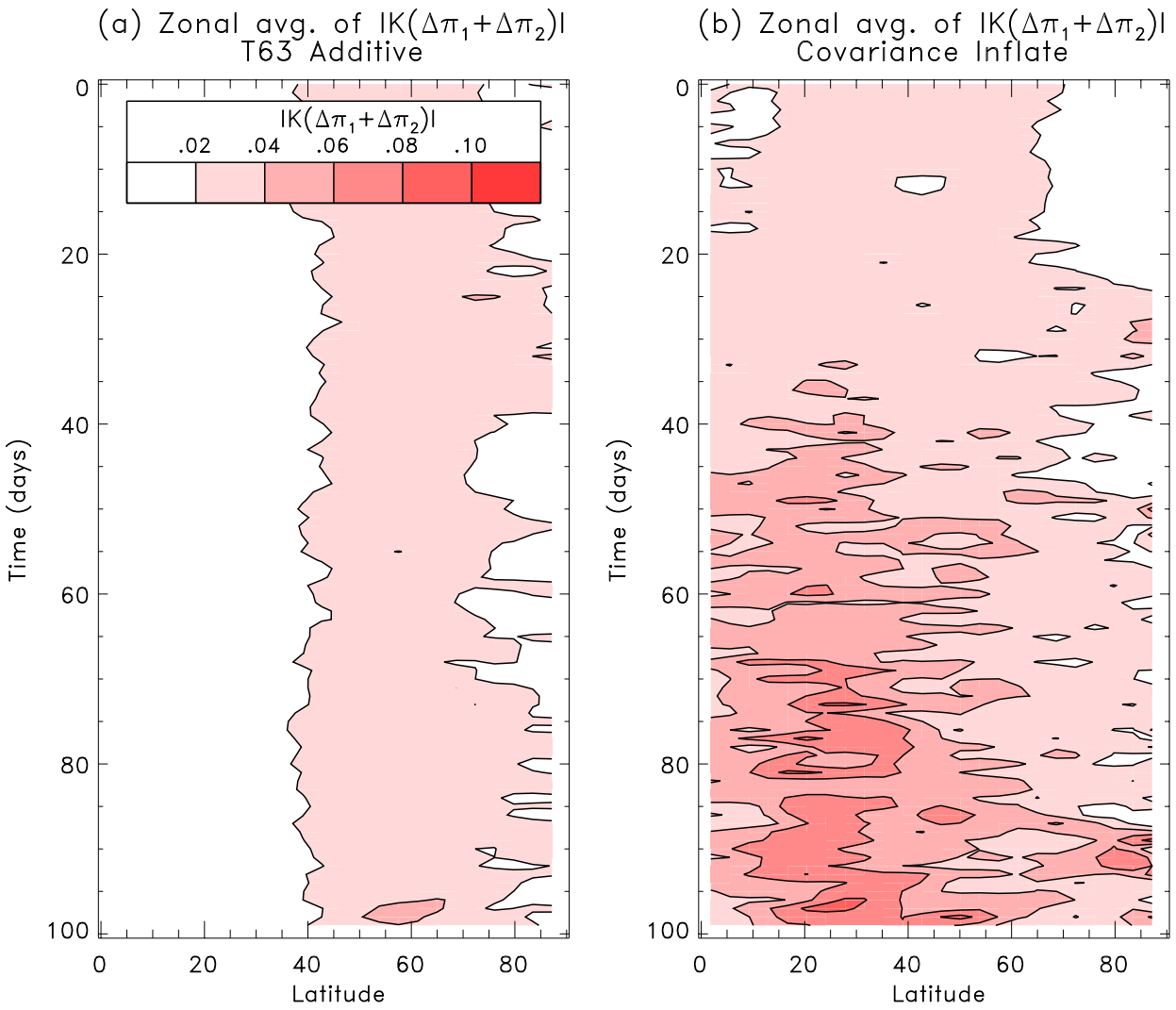

Figure 10. Hovmoller diagrams of zonally averaged Kalman gain **K** of changes of surface Exner function $\Delta\pi_1 + \Delta\pi_2$ from the (a) T63 additive experiment, and (b) covariance inflation experiment.